# High Performance Absorption Algorithms for Terminological Reasoning

Ming Zuo and Volker Haarslev
Concordia University, Montreal, Quebec, Canada
{ming_zuo|haarslev}@cse.concordia.ca

**Abstract**

When reasoning with description logic (DL) knowledge bases (KBs), performance is of critical concern in real applications, especially when these KBs contain a large number of axioms. To improve the performance, axiom absorption has been proven to be one of the most effective optimization techniques. The well-known algorithms for axiom absorption, however, still heavily depend on the order and the format of the axioms occurring in KBs. In addition, in many cases, there exist some restrictions in these algorithms which prevent axioms from being absorbed. The design of absorption algorithms for optimal reasoning is still an open problem. In this paper, we propose some new algorithms to absorb axioms in a KB to improve the reasoning performance. The experimental tests we conducted are mostly based on synthetic benchmarks derived from common cases found in real KBs. The experimental evaluation demonstrates a significant runtime improvement.

## 1 Motivation

When reasoning with description logic (DL) knowledge bases (KBs) which contain a large number of axioms, performance is the key concern in real application. To improve the reasoning performance, many optimization algorithms and techniques are employed in most of the modern reasoners such as RACER and FaCT++. Among the optimization algorithms, lazy unfolding is proven to be one of the most effective techniques [7]. Unfortunately, lazy unfolding does not work well for KBs containing a significant number of nonabsorbable GCIs (General Concept Inclusions). A GCI is called *nonabsorbable* if it cannot be rewritten into a rule axiom. We use the term *rule axiom* to represent axioms of the form of $A \Rightarrow C$ where $A \in NC$ and $C$ is an arbitrary concept, while the form $C \sqsubseteq D$ represents a GCI where $C$ and $D$ are arbitrary concepts. The difference between $A \Rightarrow C$ and $A \sqsubseteq C$ is that $A \Rightarrow C$ represents only "if $A$ then

$C$", while $A \sqsubseteq C$ represents both "if $A$ then $C$" and "if $\neg C$ then $\neg A$" during reasoning. To convert a GCI into rule axiom(s), a technique called *absorption* is employed. Although preliminary absorption algorithms have been discussed in [6], there is little concern about the "best" absorption for optimal reasoning. The known algorithms also still heavily depend on the order and the format of axioms found in the KB of interest. In addition, in many cases, some restrictions in these algorithms prevent axioms from being absorbed [3].

## 2 Tableau Algorithm and Lazy Unfolding

Reasoning about a DL based KB is usually reduced to concept reasoning w.r.t. a *TBox*. For such kind of reasoning, a sound and complete tableau reasoning algorithm is usually employed. The basic idea behind a tableau algorithm is to take an input concept $C$ w.r.t. a *TBox* $\mathcal{T}$, and try to prove the satisfiability of $C$ w.r.t. $\mathcal{T}$ by constructing a model $\mathcal{I}$ of $C$ w.r.t. $\mathcal{T}$. This is done by syntactically decomposing $C$ so as to derive constraints on the structure of such a model. For example, any model of $C$ must, by definition, contain some individual $x$ such that $x$ is an element of $C^{\mathcal{I}}$, and if $C$ is of the form $\exists R.D$, then the model must also contain an individual $y$ such that $\langle x, y \rangle \in R^I$ and $y$ is an element of $D^{\mathcal{I}}$; if $D$ is non-atomic, then continuing with the decomposition of $D$ would lead to additional constraints. The construction fails if the constraints include an obvious contradiction, e.g., if some individual $z$ must be an element of both $C$ and $\neg C$ for some concept $C$ [2, 5].

The decomposition and construction are usually carried out by applying so-called tableau expansion rules as described in [2]. During the tableau expansion, disjunctions are added to the label of each node of the tableaux for each GCI (one disjunction is added for axioms of the form $\mathcal{C}_1 \sqsubseteq \mathcal{C}_2$; two disjunctions are added for axioms of the form $\mathcal{C}_1 \equiv \mathcal{C}_2$ [5]). This leads to an exponential increase in the search space as the number of nodes and axioms increases [4].

An intuitive optimization technique is *lazy unfolding* — it only unfolds concepts if required during the expansion process [1]. It has been described by the additional tableau rules in [5].

*Lazy unfolding* cannot be applied to an arbitrary axiom in a *TBox* due to the atomic concept restriction on the left-hand side of the axiom. However, we can still divide an arbitrary *TBox* $\mathcal{T}$ into two parts: the unfoldable part $\mathcal{T}_u$, to which we can apply *lazy unfolding* directly, and the general part $\mathcal{T}_g$, in which we have to perform reasoning by general tableau expansion [5]. Therefore, there is an intuitive optimization technique to be considered: converting general axioms from $\mathcal{T}_g$ to $\mathcal{T}_u$ while keeping the semantics of *TBox* unchanged. This is the original idea of an *"absorption"*.

## 3 Standard Absorption

Let us first consider an absorption example. Suppose we have two *TBoxes* $\mathcal{T}$ and $\mathcal{T}'$.

$\mathcal{T} = \mathcal{T}_u \cup \mathcal{T}_g$ and $\mathcal{T}_u = \emptyset$; $\mathcal{T}_g = \{A \sqsubseteq C;\ \neg A \sqsubseteq D\}$;
$\mathcal{T}' = \mathcal{T}'_u \cup \mathcal{T}'_g$ and $\mathcal{T}'_u = \{A \Rightarrow C;\ \neg A \Rightarrow D\}$; $\mathcal{T}^{'}_g = \emptyset$.

An obvious question is whether $\mathcal{T}' \equiv \mathcal{T}$?

According to the definition of a *correct absorption* proposed in [6], $\mathcal{T}'$ is a correct absorption of $\mathcal{T}$. Unfortunately, $\mathcal{T}' \neq \mathcal{T}$ since $\mathcal{T}' \models \mathcal{T}$ does not hold. To evaluate the correctness of an absorption, we introduce the notion of a *valid absorption*.

**Definition 3.1 (valid absorption)** *Let $\mathcal{T}$ be a TBox, and $\mathcal{T}'$ be an absorption of $\mathcal{T}$. If $\mathcal{T}' \models \mathcal{T}$ and $\mathcal{T} \models \mathcal{T}'$, then $\mathcal{T}'$ is called a valid absorption of $\mathcal{T}$.*

Based on the above definitions, the following absorptions are all *valid absorptions*, provided that $A$ is an atomic concept and $C, D$ and $E$ are arbitrary concepts, and $\mathcal{T}$ is an *acyclic TBox* [2].

**Proposition 3.1** *Let $\mathcal{T} = \mathcal{T}_u \cup \mathcal{T}_g$, $\mathcal{T}_u = \emptyset$ and $\mathcal{T}_g = \{A \sqsubseteq D\}$, $A \in NC$, and $\mathcal{T}' = \mathcal{T}'_u \cup \mathcal{T}'_g$; $\mathcal{T}'_u = \{A \Rightarrow D\}$ and $\mathcal{T}'_g = \emptyset$. Then $\mathcal{T}'$ is a valid absorption of $\mathcal{T}$.*

**Proposition 3.2** *Let $\mathcal{T} = \mathcal{T}_u \cup \mathcal{T}_g$, $\mathcal{T}_u = \emptyset$ and $\mathcal{T}_g = \{A \equiv D\}$, $A \in NC$, and $\mathcal{T}' = \mathcal{T}'_u \cup \mathcal{T}'_g$, $\mathcal{T}'_u = \{A \Rightarrow D; \neg A \Rightarrow \neg D\}$ and $\mathcal{T}'_g = \emptyset$. Then $\mathcal{T}'$ is a valid absorption of $\mathcal{T}$.*

**Proposition 3.3** *Let $\mathcal{T} = \mathcal{T}_u \cup \mathcal{T}_g$.*

*(1) If $\mathcal{T}'$ is an arbitrary TBox, then $(\mathcal{T}_u, \mathcal{T}_g \cup \mathcal{T}')$ is a valid absorption of $\mathcal{T} \cup \mathcal{T}'$.*

*(2) If $\mathcal{T}'$ is a TBox that consists entirely of axioms in the form of $A \sqsubseteq D$, where $A \in NC$ and neither $A$ nor $\neg A$ occur on the left-hand side in $\mathcal{T}_u$, then $(\mathcal{T}_u \cup \{A \Rightarrow D\}, \mathcal{T}_g)$ is a valid absorption of $\mathcal{T} \cup \mathcal{T}'$.*

A question rises from Proposition 3.3 whether it is possible to absorb an axiom into $\mathcal{T}_u$ if either A or $\neg A$ occur on the left-hand side of $\mathcal{T}_u$.

**Lemma 3.1** *Let $(\mathcal{T}_u, \mathcal{T}_g)$ be a valid absorption of a TBox $\mathcal{T}$. If $\mathcal{T}'$ is a TBox that consists entirely of axioms in the form of $A \sqsubseteq D$, where $A \in NC$ and $A$ already has a rule definition in $\mathcal{T}_u$, say $A \Rightarrow C$, if $\neg A$ does not appear on the left-hand side of $\mathcal{T}_u$, then $(\mathcal{T}_u \cup \{A \Rightarrow (D \sqcap C)\}, \mathcal{T}_g)$ is a valid absorption of $\mathcal{T} \cup \mathcal{T}'$.*

**Lemma 3.2** *Let $(\mathcal{T}_u, \mathcal{T}_g)$ be a valid absorption of a TBox $\mathcal{T}$. If $\mathcal{T}'$ is a TBox that consists entirely of axioms of the form $A \sqsubseteq D$, where $A \in NC$ and $\neg A$ already has a rule definition in $\mathcal{T}_u$, say $\neg A \Rightarrow C$, if $A$ does not appear on the left-hand side of $\mathcal{T}_u$, then $(\mathcal{T}_u \cup \{A \Rightarrow D\}, \mathcal{T}_g \cup \{\top \sqsubseteq C \sqcup D\})$ is a valid absorption of $\mathcal{T} \cup \mathcal{T}'$.*

**Lemma 3.3** *Let $(\mathcal{T}_u, \mathcal{T}_g)$ be a valid absorption of a TBox $\mathcal{T}$, if $\mathcal{T}'$ is a TBox that consists entirely of axioms of the form $A \sqsubseteq E$, where $A \in NC$. If both $A$*

*and $\neg A$ have a rule definition in $\mathcal{T}_u$, say $A \Rightarrow C$ and $\neg A \Rightarrow D$, then $(\mathcal{T}_u \cup \{A \Rightarrow (C \sqcap E)\}, \mathcal{T}_g \cup \{\top \sqsubseteq D \sqcup E\})$ is a valid absorption of $\mathcal{T} \cup \mathcal{T}'$.*

The above mentioned propositions and lemmas also hold for a cyclic *TBox* as long as no right-hand side concept in $\mathcal{T}_u$ *directly uses* [2] an atomic concept (regardless of a negation sign) occurring in the left-hand side of the same axiom.

## 4 Heuristic Absorption

Experimental experience suggests that reasoning efficiency is improved by either reducing the number of axioms in $\mathcal{T}_g$ or reducing the number of axioms in $\mathcal{T}_u$. Thus, we propose the "best" absorption is the one which can absorb a maximal number of axioms from $\mathcal{T}_g$ and keep only a minimal number of axioms in $\mathcal{T}_u$.

A first possible way of reducing axioms in $\mathcal{T}_g$ and $\mathcal{T}_u$ is to convert a *primitive definition* into a *complete definition*, i.e., selecting a concept definition instead of a concept inclusion by checking all axioms in the specified *TBox*. This can be achieved as follows.

Given an arbitrary TBox $\mathcal{T}$. Suppose $A$ is an atomic concept.

1. *Simplify [8] and normalize $\mathcal{T}$. After normalization, $\mathcal{T}_u = \emptyset$ and $\mathcal{T}_g$ contains a set of axioms in the form $\top \sqsubseteq C \sqcup D$ where $C$ and $D$ are either atomic concepts or role concepts. Then each axiom in $\mathcal{T}_g$ can be expressed as a set. For example, the set $\boldsymbol{G} = \{C, D\}$ represents the axiom in the form $\top \sqsubseteq C \sqcup D$. As a consequence, $\neg\boldsymbol{G}$ represents a set containing the concept $\neg C \sqcap \neg D$. Therefore, each axiom in $\mathcal{T}_g$ only contains $A$ or $\neg A$ or none of them. We also need a function* con *which returns for a given set $\boldsymbol{G}$ its represented concept, e.g., if $\boldsymbol{G} = \{C, D\}$, then $\mathrm{con}(\boldsymbol{G})$ returns $C \sqcup D$.*

2. *Initialize two sets $\mathcal{T}_{g_1}$, $\mathcal{T}_{g_2}$ to be empty and consider $A$ the chosen (fixed) atomic concept.*

3. *For any set $\boldsymbol{G}$, if $A \in \boldsymbol{G}$, then remove $\boldsymbol{G}$ from $\mathcal{T}_g$ and add an item $\neg(\boldsymbol{G}\backslash\{A\})$ to $\mathcal{T}_{g_1}$; if $\neg A \in \boldsymbol{G}$, then remove $\boldsymbol{G}$ from $\mathcal{T}_g$ and add an item $\boldsymbol{G}\backslash\{\neg A\}$ to $\mathcal{T}_{g_2}$; otherwise, keep $\boldsymbol{G}$ in $\mathcal{T}_g$.*

4. *For each item $\boldsymbol{G}_2$ in $\mathcal{T}_{g_2}$, if $\boldsymbol{G}_2$ also appears in $\mathcal{T}_{g_1}$,*
   (a) *remove $\boldsymbol{G}_2$ from both $\mathcal{T}_{g_1}$ and $\mathcal{T}_{g_2}$;*
   (b) *add the axiom $\{A \Rightarrow \mathrm{con}(\mathbf{G_2})\}$ and $\{\neg A \Rightarrow \mathrm{con}(\neg\mathbf{G_2})\}$ to $\mathcal{T}_u$.*

5. *For each item set $\boldsymbol{G}_1'$ left in $\mathcal{T}_{g_1}$, create a new set $\neg\boldsymbol{G}_1' \cup \{A\}$ and put it back into $\mathcal{T}_g$; for each item set $\boldsymbol{G}_2'$ in $\mathcal{T}_{g_2}$, create a new set $\boldsymbol{G}_2' \cup \{\neg A\}$ and put it back into $\mathcal{T}_g$.*

The above algorithm can be further improved in step (4) by checking more than one item set in $G_2$. For more general KBs, the following heuristic absorption procedure could be applied.

Given an arbitrary TBox $\mathcal{T} = \mathcal{T}_u \cup \mathcal{T}_g$. $\mathcal{T}_u$ contains a set of axioms of the form $C \equiv D$ or $C \Rightarrow D$. $\mathcal{T}_g$ contains a set of axioms of the form $\top \sqsubseteq C$.

1. *Simplify and normalize $\mathcal{T}$ as described above.*

2. *Suppose $c_p$ is the total number of appearance of $A$ in $\mathcal{T}_g$; and $c_n$ is the total number of appearance of $\neg A$ in $\mathcal{T}_g$. Among all $c_p$s and $c_n$s, select the greatest value and absorb the accordingly atomic concept to $\mathcal{T}_u$.*

3. *Repeat step (2) until no concept can be absorbed anymore.*

In fact, the above mentioned two algorithms can be combined to achieve a better performance.

## 5 Experimental Results

To check the effectiveness of the newly developed algorithms, we compare them with the ones currently employed by RACER. Firstly, we developed a customized RACER version by disabling its absorption module. After that, we developed a new external absorption module (as a stand-alone program) by implementing the algorithms described above. Each of our test KBs is processed by the external absorption module. Its output is used as input to the customized RACER version. We also process each original test KB with the standard RACER version. In our graphs we compare for each test KB the TBox classification time of the customized RACER version (denoted as "enhanced") with the standard version of RACER (denoted as "normal").

### 5.1 Primitive Definition to Complete Definition

Suppose $A, B, C, D$ are all atomic concepts in the following example:

*TBox* $\mathcal{T} = \mathcal{T}_u \cup \mathcal{T}_g$;
$\mathcal{T}_u = \{A \Rightarrow B;\ C \Rightarrow B;\ B \Rightarrow A \sqcup C\}$ and $\mathcal{T}_g = \{\top \sqsubseteq A \sqcup C \sqcup \exists R.D\}$

After normalization and simplification, the above *TBox* can be easily absorbed into the following *TBox* $\mathcal{T}'$ by heuristically creating these concept definitions:

$\mathcal{T}_u' = \{B \equiv (A \sqcup C); \neg A \Rightarrow C \sqcup \exists R.D\}$
$\mathcal{T}_g' = \emptyset$

The classification times for a KB where we replicated this pattern of axiom pairs are shown in the left graph in Figure 1.

From the graph one can see that the CPU time for reasoning with the algorithms currently employed by RACER is exponential. However, the performance based on the newly developed absorption algorithm is roughly linear.

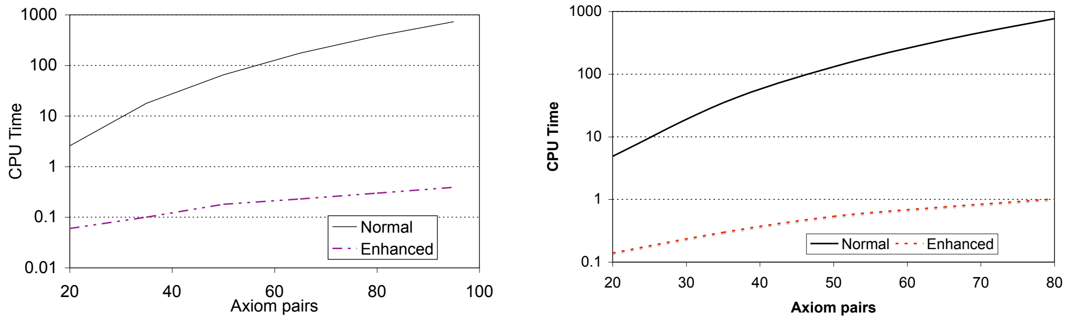


Figure 1: Classification times (seconds) for pattern "primitive to complete definition" (left graph) and for pattern "enhanced absorption" (right graph).

### 5.2 Enhanced Absorption Algorithm

Based on Lemma 3.2 and 3.3, we were able to develop an absorption algorithm by allowing both positive or negative atomic concepts to occur on the left-hand side of $\mathcal{T}_u$. Consider the following example ($A, B \in NC$):

$$\mathcal{T} = \mathcal{T}_u \cup \mathcal{T}_g;$$
$$\mathcal{T}_u = \{A \sqsubseteq B\};\ \mathcal{T}_g = \{\top \sqsubseteq A \sqcup \exists R.K\}$$

By applying Lemma 3.3, we are able to completely absorb it as the following:

$$\mathcal{T}_u = \{A \Rightarrow B; \neg A \Rightarrow \exists R.K; \neg B \Rightarrow \exists R.K\};\ \mathcal{T}_g = \emptyset$$

The test result is shown in the right graph in Figure 1. They show that the new absorption algorithm is much more effective than the classical absorption algorithm due to the elimination of absorption restrictions.

### 5.3 Heuristic Absorption

Suppose we have the following *TBox* $\mathcal{T}$ which consists of the following general axioms:

$$\{\neg A \sqcup B;\ \neg C \sqcup \neg D \sqcup \exists R_1.C_1 \sqcup \exists R_2.C_2;\ \neg D \sqcup K;\ \neg M \sqcup N;\ C \sqcup D \sqcup \exists R_1.C_1 \sqcup M \sqcup A\}$$

To absorb $\mathcal{T}$, we follow the procedure described in Section 4:

**Step 1**: List the statistics of $T_g$ for each atomic concept. The result is as follows. (The format is $A(c_p, c_n)$, where $c_p$ gives the number of positive (unnegated) occurrences and $c_n$ the number of negative (negated) occurrences, and we ignore the atomic concepts occurring in the qualifications of existential and universal restrictions.)

A(0,1); B(1,0);C(1,1);D(1,2);M(1,1); N(1,0); K(1,0);$C_2(0,0)$; $C_1(0,0)$

We divide the atomic concepts into two groups by selecting the concepts where one of $c_p$ or $c_n$ values is zero and discarding the concepts that have both $c_p$ and $c_n$ values equal to zero. We obtain the following:

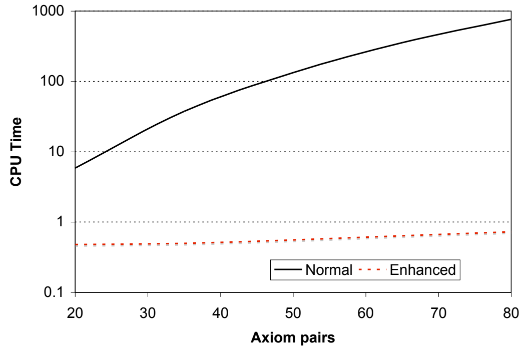


Figure 2: Classification times (seconds) for pattern “heuristic absorption”.

Group 1: B(1,0); N(1,0); K(1,0)
Group 2: A(1,1); C(1,1);D(1,2);M(1,1)

**Step 2**: We give the concepts in the first group a higher priority. Then, the concepts B, N, K have the same priority. Suppose we absorb B first. We have $T_u$ and $T_g$ as the following:

$T_u$: $\{\neg B \Rightarrow \neg A\}$
$T_g$: $\{\neg C \sqcup \neg D \sqcup \exists R_1.C_1 \sqcup R_2.C_2;\ \neg D \sqcup K;\ \neg M \sqcup N;\ C \sqcup D \sqcup \exists R_1.C_1 \sqcup M \sqcup A\ \}$

Statistics group:
Group 1: A(1,0); N(1,0); K(1,0)
Group 2: C(1,1);D(1,2);M(1,1)

We repeat step 1 and step 2 until $T_g$ is empty. At the end, $T_u$ contains the following axioms:

$\neg B \Rightarrow \neg A$
$\neg A \Rightarrow C \sqcup D \sqcup \exists R_1.C_1 \sqcup M$
$D \Rightarrow K \sqcap (\neg C \sqcup \exists R_1.C_1 \sqcup R_2.C_2)$
$M \Rightarrow N$

and $T_g = \emptyset$

The test results using the above absorption scheme are shown in Figure 2. The improvement of the heuristic absorption algorithm is significant compared with the one currently employed by RACER.

## 6 Conclusion

We proposed criteria for the “best” absorption based on experimental experience. Then, we introduced novel heuristic absorption algorithms. We have demonstrated how these algorithms are working, and how they affect the reasoning performance. In addition, we have shown that some restrictions applied in the absorption algorithms of RACER could be eliminated. Therefore, the absorption algorithms can be effectively applied to more general axioms.

We have also implemented the heuristic algorithm by incorporating the optimizations known from the RACER reasoner. We have illustrated their effectiveness by analyzing the reasoning performance of RACER when classifying benchmark KBs. The analysis shows that, not only are the new techniques highly effective, but also the reasoning performance is not significantly affected by the order and format of the axioms occurring in a KB.

## References

[1] F. Baader, E. Franconi, B. Hollunder, B. Nebel, and H. Profitlich. An empirical analysis of optimization techniques for terminological representation systems. In *Proc. of the 3rd Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR92)*, pages 270–281, 1992.

[2] F. Baader and W. Nutt. Basic description logics. In F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, and P.F. Patel-Schneider, editors, *The Description Logic Handbook: Theory, implementation, and applications*, pages 47–100. 2003.

[3] V. Haarslev and R. Möller. High performance reasoning with very large knowledge bases. In *Proceedings of Seventeenth International Joint Conference on Artificial Intelligence*, pages 161–166, 2001.

[4] I. Horrocks. *Optimising Tableaux Decision Procedures for Description Logics.* PhD thesis, University of Manchester, 1997.

[5] I. Horrocks. Implementation and optimisation techniques. In F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, and P.F. Patel-Schneider, editors, *The Description Logic Handbook: Theory, implementation, and applications*, pages 313–355. 2003.

[6] I. Horrocks and S. Tobies. Optimisation of terminological reasoning. In *Proc. of the 2000 Description Logic Workshop*, pages 183–192, 2000.

[7] I. Horrocks and S. Tobies. Reasoning with axioms: Theory and practice. In *Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning*, pages 285–296, 2000.

[8] F. Massacci. Simplification: A general constraint propagation technique for propositional and modal tableaux. In *Proceedings of the International Conference on Automated Reasoning with Analytic Tableaux and Related Methods*, pages 217–231, 1998.